

AUTOMATED DISEASE-DRUG ONTOLOGY GENERATION FRAMEWORK POWERED BY LINKED BIOMEDICAL ONTOLOGIES

The Oakland University and School of Engineering and Computer Science communities are invited to attend Mazen Alobaidi's defense of his Ph.D. dissertation. Seating is limited. RSVP with Katie Loodeen at loodeen@oakland.edu.

Automated Disease-Drug Ontology Generation Framework Powered By Linked Biomedical Ontologies

Committee: Khalid Mahmood Malik, Ph.D. (Chair), Huirong Fu, Ph.D., Erik Fredricks, Ph.D., Mohammad Dadashzadeh, Ph.D.

Time: 12:00 – 2:00 p.m.
Date: Wednesday, Feb 13, 2019
Location: 347 EC

The exponential growth of the unstructured data available in biomedical literature, and Electronic Health Record (EHR), requires powerful novel technologies and architectures to unlock the information hidden in the unstructured data. The success of smart healthcare applications such as clinical decision support systems, disease diagnosis systems, and healthcare management systems depends on knowledge that is understandable by machines to interpret and infer new knowledge from it. In this regard, ontological data models are expected to play a vital role to organize, integrate, and make informative inferences with the knowledge implicit in that unstructured data and represent the resultant knowledge in a form that machines can understand. However, constructing such models is challenging because they demand intensive labor, domain experts, and ontology engineers. Such requirements impose a limit on the scale or scope of ontological data models. This thesis presents a framework that allows mitigating the time-intensity to build ontologies and achieve machine interoperability by taking unstructured data as input from published PubMed abstracts. Empowered by linked biomedical ontologies, proposed novel Automated Ontology Generation Framework consists of five major modules: a) Text Processing using compute on demand approach. b) Medical Semantic Annotation using N-Gram, ontology linking and classification algorithms, c) Relation Extraction using graph method and Syntactic Patterns, d) Semantic Enrichment using RDF mining, e) Domain Inference Engine to build the formal ontology. Quantitative evaluations show 84.78% recall, 53.35% precision, and 67.70% F-measure in terms of concepts identification; 85.51% recall, 69.61% precision, and F-measure 76.74% with respect to taxonomic relation extraction; and 77.20% recall, 40.10 % precision, and F-measure 52.78 % with respect to biomedical non-taxonomic relation extraction. This dissertation concludes that proposed natural language processing, semantic enrichment, syntactic pattern, and graph algorithm based techniques along with use of Linked Biomedical Ontologies is promising solution to the problem of automating the process of disease-drug ontology generation.

